

# Welcoming More Participation in Open Data Science for the Oceans

Alexa L. Fredston<sup>1</sup> and Julia S. Stewart Lowndes<sup>2</sup>

<sup>1</sup>Department of Ocean Sciences, University of California, Santa Cruz, California, USA;  
email: fredston@ucsc.edu

<sup>2</sup>National Center for Ecological Analysis and Synthesis, University of California, Santa Barbara,  
California, USA

Annu. Rev. Mar. Sci. 2024. 16:537–49

First published as a Review in Advance on  
July 7, 2023

The *Annual Review of Marine Science* is online at  
[marine.annualreviews.org](http://marine.annualreviews.org)

<https://doi.org/10.1146/annurev-marine-041723-094741>

Copyright © 2024 by the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.

## Keywords

open science, data science, inclusion

## Abstract

Open science is a global movement happening across all research fields. Enabled by technology and the open web, it builds on years of efforts by individuals, grassroots organizations, institutions, and agencies. The goal is to share knowledge and broaden participation in science, from early ideation to making research outputs openly accessible to all (open access). With an emphasis on transparency and collaboration, the open science movement dovetails with efforts to increase diversity, equity, inclusion, and belonging in science and society. The US Biden–Harris Administration and many other US government agencies have declared 2023 the Year of Open Science, providing a great opportunity to boost participation in open science for the oceans. For researchers day-to-day, open science is a critical piece of modern analytical workflows with increasing amounts of data. Therefore, we focus this article on open data science—the tooling and people enabling reproducible, transparent, inclusive practices for data-intensive research—and its intersection with the marine sciences. We discuss the state of various dimensions of open science and argue that technical advancements have outpaced our field’s culture change to incorporate them. Increasing inclusivity and technical skill building are interlinked and must be prioritized within the marine science community to find collaborative solutions for responding to climate change and other threats to marine biodiversity and society.

ANNUAL  
REVIEWS **CONNECT**

[www.annualreviews.org](http://www.annualreviews.org)

- Download figures
- Navigate cited references
- Keyword search
- Explore related articles
- Share via email or social media

OPEN  ACCESS 

## 1. WELCOME TO MARINE OPEN DATA SCIENCE

Open science has many different definitions, and we can participate in a spectrum of products (e.g., data and publications) and practices (e.g., sharing ideas and work in progress). One definition is that open science is “a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding” (Ramachandran et al. 2021, p. 2). The US Biden–Harris Administration, the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Administration (NOAA), the National Science Foundation, and dozens of other US government agencies have declared 2023 to be the Year of Open Science (NASA 2022, White House Off. Sci. Technol. Policy 2023) and are developing strategic policy to promote equitable participation. Focusing more specifically on data analytical workflows and sharing in the marine sciences, we define open data science as “the tooling and people enabling reproducible, transparent, inclusive practices for data-intensive research” (Lowndes & Robinson 2022). Open science is often interpreted to be synonymous with open data as a product; our definition of open data science goes far beyond data sharing and includes the social and technical processes that researchers encounter when working with data.

Open data science tools and practices are revolutionizing science by enabling transparent, collaborative, and reproducible data-driven research, with recent examples including the use of openly shared data to facilitate real-time decision-making amid the COVID-19 pandemic (Zastrow 2020) and the use of open source software to capture the first images of black holes (NumFOCUS 2023). Most known for underpinning robust data analyses and visualization, open data science also streamlines collaboration and expands communications through modern online channels for contributing to, publishing, and distributing research outputs (Bastille et al. 2021, Lowndes et al. 2017, McKiernan et al. 2016). Aiming to make these practices the norm, scientific journals are increasingly requiring that authors submit all data and code used for the analyses they publish (Berberi & Roche 2022). Simultaneously, many researchers who may not identify as data scientists—including marine scientists—are working with larger and more complex datasets than ever before, and code is now a requirement to do their science (Geiger 2018, Stoudt et al. 2021). Open access to these datasets enabled the data synthesis revolution and now allows scientists to address foundational questions in fields like the marine sciences at scales that were previously unimaginable (Halpern et al. 2020). Yet in the marine sciences, large gaps remain between best practices in open science and the status quo (Hörstmann et al. 2021, Lowndes et al. 2017), and a lack of data is still a common problem (Blasco et al. 2020). While the open science movement is widespread in research generally, we identify two reasons why it is useful to explore its intersection with marine science specifically.

First, the marine sciences are awash in far more data than ever before. For example, the *Tara* Oceans Consortium has published molecular and environmental data from more than 35,000 at-sea samples (Pesant et al. 2015), and the Argo Program has shared temperature and salinity data from more than 2 million vertical profiles (Wong et al. 2020). Yet what each of us hears from our peers and networks is that researchers do not know what data to use, where to find them, and how to analyze them properly. We argue that marine science is not lacking in technical tools but cannot fully realize the potential of new data streams and computational methods without systemic changes in our teams, communities, and institutions that underpin modern analytical skills and collaborative culture.

Second, the geosciences remain one of the least diverse STEM fields (Burton et al. 2023, Marín-Spiotta et al. 2020), along with ecology (Primack et al. 2023). We recognize that these inequities, and potential solutions, are highly varied and multidimensional. We believe—based in part on our lived experiences—that open data science can promote a sense of community

and belonging, thereby helping to build a more inclusive marine science (Fenwick et al. 2023, Gaynor et al. 2022, Lowndes 2019). In other words, filling the gaps in social infrastructure for open science—via skill building, culture change, and work–life balance—can also move us toward more diverse, equitable, and inclusive marine sciences.

Open data science is a continuum and not all-or-nothing; anywhere we are starting from is the right place, and we can participate anywhere along this continuum, from open hypotheses to open data and code to open publishing. Open data science breaks the narrative of “I work alone” and instead develops a mindset of reuse and of collaboration over competition, where we learn with, from, and for others (Lowndes 2019). This is a mindset we can develop and role-model for others, in any part of our careers. One way to participate incrementally is to work openly with ourselves and our teams, sharing ideas, slides, and other work in progress, as well as noticing who is participating, asking questions, listening, and creating inclusive spaces for others to voice ideas and share as well (Robinson & Lowndes 2022). Here, we frame our vision for the future of marine science around open data, open code, and open publishing and share our personal narratives of what it looks like to be a modern marine researcher practicing open science.

## 2. INTRODUCING OURSELVES: BUILDING CAREERS WITH OPEN SCIENCE

As a marine scientist starting a PhD in 2014, Alexa Fredston experienced and was influenced by open science from the very beginning, although she did not realize it at the time. Her early research questions focused on understanding climate-related shifts in marine species’ geographical ranges, questions she knew were possible to answer by combining data from bottom-trawl surveys and hindcast oceanographic data products. While the concepts were clear, the skills and tools needed to analyze the data were far from obvious. Fredston learned with Julia Lowndes and the Ocean Health Index team at the National Center for Ecological Analysis and Synthesis (NCEAS) about coding collaboratively in R, shared version control with GitHub, and open educational resources like online tutorials (Lowndes et al. 2017). EcoDataScience (<https://eco-data-science.github.io>), a local community of practice led by the Ocean Health Index team, gave her a supportive and encouraging space in which to learn—and soon teach—crucial data science techniques. This PhD experience was somewhat unique in that she was trained in open source technical skills as well as a culture of reuse, continued learning, sharing, and inclusive collaboration from the start. Since then, she has published manuscripts as preprints, shared data and code openly throughout her research workflow, participated in an international collaboration aimed at data harmonization (Maureaud et al. 2023), and joined the Board of Directors of the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology (SORTEE; <https://sortee.org>).

Meanwhile, open science was also changing Lowndes’s career path. Open data science enabled the Ocean Health Index team at NCEAS to work more reproducibly and efficiently, building from an established team culture of trust and horizontal leadership (Lowndes et al. 2017). The team’s work was enabled by interacting more with open source software developers and community builders through groups such as rOpenSci (<https://ropensci.org>), Posit (formerly RStudio; <https://posit.co>), R-Ladies (<https://rladies.org>), the Carpentries (<https://carpentries.org>), and Mozilla (<https://mozilla.org>), groups intentionally building a culture for kinder science (Lowndes 2019).

Feeling she could contribute to science most impactfully by supporting other researchers rather than continuing research herself, in 2018 Lowndes founded Openscapes (<https://openscapes.org>) through a Mozilla Open Science Fellowship and began formally mentoring marine science teams in academia (Lowndes et al. 2019) and NOAA Fisheries. In 2020, she partnered with Erin Robinson, a leader in the NASA Earthdata community, and began working with NASA

Earthdata teams (Lowndes & Robinson 2021) and communities including the Earth Science Information Partners (ESIP), the International Interactive Computing Collaboration (2i2c), Pangeo, pyOpenSci, NASA Transform to Open Science (TOPS), Ladies of Landsat, Code for Science and Society, Black in Marine Science (BIMS), and Black Women in Marine Ecology, Evolution, and Marine Biology (BWEEMS). Her collaborative work stems from a “kinder science for Future Us” mindset to welcome and empower marine, environmental, and Earth scientists with existing open source tooling and practices, helping them develop collaborative skills and join existing communities in the open science movement.

Stories like ours are important because they show a few of the many ways to participate in open science and show what open science can look in the Year of Open Science. We acknowledge that we are two white women who felt safe sharing our work openly and were supported in our jobs to learn the skills needed to practice open science. Safety and support are two elements that are fundamental to increasing participation in open science in marine ecology, and something for which we continue to fiercely advocate.

### 3. REUSING AND CONTRIBUTING DATA

Open data are freely available for reuse online in many data repositories (see table 1 in Culina et al. 2018). Those repositories are how many of us as marine scientists first engage in open science. For example, we use sea surface temperature data from NOAA and NASA satellites and find data published with a particular study in data repositories like Dryad and the Open Science Framework (OSF). We are able to reuse and cite these data as we would a research article to give proper attribution. Then, when we share our own data to these repositories, they are assigned a digital object identifier (DOI) that others can use to credit us. Metadata—data that describe and give information about other data, such as where data were collected and descriptions of the column headers—are critical for both searching for existing data and contributing new data. Guidance is available for how and where to share data, including how to format metadata and consider FAIR (findable, accessible, interoperable, and reusable) and CARE (collective benefit, authority to control, responsibility, and ethics) principles (Barker et al. 2022, Carroll et al. 2020, Roche et al. 2022b).

However, finding these datasets can be challenging. From genes to populations to spatial data products, our field’s synthesis and publication of data have vastly outpaced our guidance for their interpretation and reuse. Among the numerous calls to share and store environmental data so they are not lost (Bledsoe et al. 2022, Poisot et al. 2019, Wolkovich et al. 2012), open datasets have proliferated, and so have places to host them. One common question that we, as marine synthesis scientists, often hear is, Which data source should I use? While it is encouraging that it is becoming more straightforward to upload and share data, this leads to difficulty in selecting and interpreting datasets for reuse. For example, despite the relative ease of finding a dataset on ocean temperature, there is almost no guidance to help marine scientists decide which ocean temperature dataset to use and why.

This challenge is far beyond what we can address in this review, but we advocate for working groups and other interdisciplinary collaborations to not only publish frequently updated guidelines to available datasets in their fields—pros, cons, and when to use each one—but help teach others to normalize the practice. Important road maps for datasets exist, for example, through the Arctic Data Center (<https://arcticdata.io>), in ecology and evolution (Culina et al. 2018), and in marine environmental DNA (eDNA) metabarcoding (Shea et al. 2023), emphasizing the need to publish and maintain such road maps. Regardless of the topic, each set of guidelines will likely need to address how to decide among seemingly similar datasets as well as questions about spatial and temporal scale, uncertainty, and reconciling contradictory data points.

Furthermore, accessing and working with data require technical skills that may or may not be familiar to researchers. Datasets that follow FAIR guidelines have lower barriers to reuse because they are designed to be found by a broad audience and reused for various applications. Field-specific standards to enable interoperability exist or are under development in many disciplines, such as the Darwin Core standards for biodiversity data (Darwin Core Task Group 2009) and Minimum Information About Any Sequence (MIxS) for genomic data (Genome Stand. Consort. 2023). While field-specific standards are a significant step, accurately interpreting and reusing these datasets nonetheless require domain expertise that marine scientists researching ever-broader topics may not have. Sharing each field's best practices for common operations, such as downscaling spatial data layers or extrapolating from traits of closely related species, is critical along with providing the data, as is support for coding and learning new software tools. These best practices would go beyond metadata to chart a course for researchers to teach themselves how data should be accurately and responsibly reused. Open communities around data specifications, such as spatial operations in the R programming language (<https://r-spatial.org>) and the Zarr array data format (<http://zarr.dev>), are important for marine scientists to engage with.

#### 4. CODING AND LEARNING NEW SOFTWARE TOOLS

Coding (and associated skills such as version control, e.g., with Git) is increasingly critical for modern marine scientists, as the use of climate models, oceanographic data products, advanced statistical methods, and complex and multidimensional datasets in research projects has become routine (Braga et al. 2023, Ram 2013). Open source languages enable broader participation in coding since they are free to access and use. Because they are coproduced by an entire community, open source languages can also more nimbly evolve to match new data, techniques, and ideas. In our experience, R and Python are the primary open languages used by marine scientists, and Julia is emerging as another (additional languages include MATLAB, C++, and Fortran). Learning to code should be done with “good enough practices” (Wilson et al. 2017) and community resources rather than ad hoc or alone (Lowndes et al. 2017, McKiernan et al. 2016). This means learning to write collaborative code that we expect others to see, understand, and reuse—most importantly, for a researcher's future self (Wilson et al. 2017) long after the code was originally written, i.e., Future You and Future Us (Lowndes et al. 2019, Robinson & Lowndes 2022). Community resources relevant to marine scientists include rOpenSci and pyOpenSci, both of which also focus on code review and package development, which are important practices for creating scientific analyses via code.

Digital notebooks like Jupyter Notebooks and R Markdown help us iterate between writing, coding, data visualization, and related tasks while we work, since our scripts, outputs (figures and tables), and text are all in the same place (e.g., Czapanskiy et al. 2022, Ovando et al. 2021). Furthermore, they have changed how we approach science communication and publishing since we can create not only Word documents and PDFs but also webpages, websites, e-books, slides, and more from the same source as a part of our daily workflows. The interoperability between these tools is increasing with tools like Quarto, which enables you to publish websites that are a combination of Jupyter and R Markdown notebooks, as well as the reticulate package, which enables you to run Python code from R. These tools enable researchers to leverage tools developed in a variety of open source languages with less need to translate between them. Open documentation to guide researchers through these tools includes guidelines for best practices in archiving data, metadata, and scripts (Gil et al. 2016, Jenkins et al. 2023, Reichman et al. 2011), code writing (Filazzola & Lortie 2022, Wilson et al. 2017), and collaborative workflows (Barros et al. 2023, Lowndes et al. 2017). Furthermore, the FAIR and CARE standards can be applied to a broad range of research software (Barker et al. 2022).

Learning to code and use new software takes time, and learning is a continual process that is never done. Despite coding and data skills being a large unmet need (Barone et al. 2017), individual scientists often have to advocate to make learning part of their approved paid time. Additionally, one element preventing the formal instruction of these skills in universities is that self-taught coders often feel like they are not expert enough to teach others (Williams et al. 2019, 2023). This is changing—slowly—as there is more visibility of this skills gap and more groups that teach at a community level, such as R-Ladies and the Carpentries. It is encouraging to see more research software engineers in academia—following long-standing work by groups like the US Research Software Sustainability Institute (<https://urssi.us>) and ESIP—and we hope to see increasing numbers of research software engineers in the marine sciences collaborating in the future.

## 5. RETHINKING SCIENTIFIC PUBLISHING: SHARING EARLIER AND REWARDING MORE

Journals disseminating scientific results have been a mainstay of the research process for centuries. However, profound issues endemic to the modern academic publishing enterprise—such as the exorbitant costs of accessing articles managed by for-profit publishers and academia’s overemphasis on journal prestige—are also not new (Walter & Mullins 2019). Rather than recapitulate these well-described issues, we want to bring awareness to several new dimensions of scientific publishing that are affecting all fields, including the marine sciences.

More and more researchers are sharing their work publicly at the manuscript draft stage as preprints. A preprint is typically hosted on a dedicated server—such as EcoEvoRxiv (<https://ecoevorxiv.org>), bioRxiv (<https://biorxiv.org>), EarthArXiv (<https://eartharxiv.org>), or OSF Preprints (<https://osf.io/preprints>)—that generates a DOI for the document. Preprint servers make it easy to update manuscripts with new versions, so researchers often upload an early draft and iteratively improve it with feedback and peer review; these servers are also increasingly leveraging notebook technology (see Section 4). If the article is eventually published in a peer-reviewed journal, authors can link it to the preprint version. Citation counts for the preprint and the published article will be pooled on sites like PubMed and Google Scholar, primary websites for tracking researcher citations. Scientific journals increasingly support preprints, with some even offering an option for researchers to publish their manuscript draft as a preprint upon submission to the journal. Awareness of preprints skyrocketed during the COVID-19 pandemic as scientists raced to share data and research to understand infection rates and spread (Watson 2022). Of course, the key caveat of preprints—that they have not yet been peer-reviewed and deemed sound for publication by experts—remains.

The core function of a preprint server—to be publicly accessible and free—makes it an appealing option for practitioners of open science. Preprinting a draft manuscript allows researchers to share their ideas with the scientific community much earlier, potentially amplifying the impact of their work and increasing media attention and citation counts (Fu & Hughey 2019). For example, an article Alexa Fredston coauthored that was first preprinted on January 25, 2022 was published in a peer-reviewed journal on April 4, 2023; by that time, it already had more than 1,500 views, almost 500 downloads, and several citations and had stimulated some discussion on social media. Google Scholar merged those citations and the digital record of the preprint with the journal’s version of the article within a week of its publication. Preprints are also a primary route to green open access (in which some version of the final manuscript, such as a preprint, is freely available online; gold open access means that the publisher-formatted final version is freely available), which is key to complying with funder mandates for open access (Roche et al. 2022a).

Several journals have proposed creative models that aim to improve academic publishing. The Public Library of Science (PLOS) operates its open access-only journals as a nonprofit, and its

journal *PLOS ONE* does not evaluate submissions for novelty; the novelty criterion is one possible culprit for the bias toward statistically significant results in published literature (Fanelli 2012). Other journals publish registered reports, in which authors receive peer review and may have articles provisionally accepted after analyses have been designed but before the study is actually conducted, to minimize publication bias and promote preregistration (O’Dea et al. 2021). eLife, also a nonprofit, recently transitioned to a reviewed-preprint model where all submissions that are sent out for peer review are eventually published open access with the associated peer reviews. Journals published by academic societies, which are also typically nonprofits that offer open access publishing options, use revenue from publishing for society activities and likely provide a greater social good than strictly for-profit journals (Chytrý et al. 2023).

Whatever their model, all journals are now confronting a seismic shift toward open access in scientific publishing (Butler et al. 2023), and their policies are evolving rapidly [the Sherpa Romeo tool (<https://v2.sherpa.ac.uk/romeo>) is a way to find the policies of specific journals]. A growing awareness of the difficulty of accessing paywalled research led to Plan S, which launched in Europe in 2018 with the requirement that state-funded research be published open access by 2021; agencies from around the globe have since signed on to Plan S (Eur. Sci. Found. 2023). The US government declared in 2022 that all research done with taxpayer funds must be published open access by 2025 (White House Off. Sci. Technol. Policy 2022).

Although open access publishing removes one key barrier to equity in science—the paywall—it has been criticized for introducing another one: The article publishing charges levied by journals to publish articles gold open access typically run in the thousands of dollars and are prohibitively costly for scientists working outside of traditional research institutions in the wealthiest nations. One scholar estimated that the average article publishing charge is equivalent to half a year’s pay (or tuition and stipend, for a student) in many African nations, highlighting the inaccessibility of gold open access publishing in most of the world (Mekonnen et al. 2021). Green open access does not charge fees to authors, however, and many other models of open access publication exist. For example, diamond or platinum open access journals do not charge fees to either authors or subscribers. The initiative around Plan S (cOAlition S) and many others are working to expand diamond open access journals, which currently serve “a fine-grained variety of generally small-scale, multilingual, and multicultural scholarly communities” (Ancion et al. 2022, p. 3).

Publishing preprints and taking an inclusive approach to authorship (for a review of issues and solutions regarding authorship, see Cooke et al. 2021, Nakagawa et al. 2023) are necessary but not sufficient conditions for advancing open science in the marine sciences. The incentive and reward structure for professional marine scientists must also adapt to this new paradigm. Specifically, in assessment, tenure, and promotion, collaboration and nonpublication outputs must be rewarded more, and publishing in prestige journals must be rewarded less (Leonelli et al. 2015, Merow et al. 2023, Natl. Acad. Sci. Eng. Med. 2018, Nosek et al. 2015). Road maps to achieve this at the institutional level already exist [the Coalition for Advancing Research Assessment (<https://coara.eu>) and the Declaration on Research Assessment (<https://sfdora.org>)]. Early-career researchers still frequently receive the advice that a single first-author paper is worth 5 or 10 coauthored papers. Especially at research-intensive universities, publications—especially in prestige journals—massively outweigh other open science contributions, such as data, code, and educational materials. This mindset devalues precisely the inclusive team spirit that we believe is a vital ingredient for high-quality, data-intensive open science.

## 6. ENGAGING COMMUNITY SCIENCE FOR THE SEA

While the marine sciences have work to do in how we share, access, and reuse data, code, and manuscripts, we also acknowledge that many marine research projects remain data limited—a

gap that can be partially filled using data collected by nonprofessionals (Binley & Bennett 2023). The value of community science or citizen science to the natural sciences is exemplified by the growing body of research using data from software like iNaturalist (<https://inaturalist.org>) and eBird (<https://ebird.org>) (Binley et al. 2023). These smartphone apps are both educational, teaching users to identify the taxa around them, and scientific, allowing those users to log taxon identifications and associated metadata. Data from these apps are freely available online and have been used for an enormous range of research and conservation purposes (Callaghan et al. 2022, Sullivan et al. 2017). Because these apps are designed to help the public interpret the natural world around them, it is no surprise that the data from iNaturalist and eBird skew heavily toward terrestrial taxa. Community science has been slow in coming to the oceans, but examples do exist, like JellyWatch (<https://jellywatch.org>), Go-Sea (<https://www.inaturalist.org/projects/go-sea>), the Secchi Disk study (Secchi Disk Seafarers et al. 2017), and ocean-focused bioblitzes (intensive biodiversity surveys done in a short time by community scientists), and we suspect that explosive growth in marine science apps and community engagement is right around the corner. Some of these initiatives developed new platforms, and others leveraged existing technology like iNaturalist with a new emphasis on marine systems.

One example of how valuable community science can be for marine science is Redmap (Range Extension Database and Mapping) Australia, a tailored public outreach initiative and associated app that encourages users to record any marine organism that seems uncommon, unfamiliar, or out of place. Similar to the other apps mentioned, Redmap seeks both to collect data on range-shifting marine species and to educate the public about the effects of climate change (Pecl et al. 2019). The program has succeeded on both counts: The community scientists demonstrated more trust and social license toward fellow marine stakeholders (Kelly et al. 2019), and the Redmap dataset has been used to model fish species responding to warming oceans (Champion et al. 2018). With myriad social and scientific benefits, we hope that these types of programs receive the sustained funding and attention that they need to be implemented throughout global coastal communities.

## **7. BROADENING PARTICIPATION IN MARINE OPEN DATA SCIENCE**

Surmounting the formidable challenges facing the marine sciences—as our field strives to understand fundamental processes, conserve biodiversity, manage natural resources, and forecast future states even while cumulative human impacts to the oceans mount—requires widespread uptake of open science practices. The way that data are collected, stored, and shared; the structure and inclusivity of collaborative teams; the computational methods we use; and the pathways for communicating scientific results are all rapidly evolving as a result of the open science movement. The tools for open marine science exist and are becoming easier to learn and more accessible. The main barriers to uptake of open science in the marine sciences, as in some other fields (Hipsley & Sherratt 2019), are often social challenges: figuring out which datasets to reuse and how to do so correctly, having supported time to learn new technical skills, navigating the incentive structure of academic publishing, and reforming institutions to encourage open science practices.

It can be overwhelming to realize that the marine science field now expects us to utilize cutting-edge software tools and meet ever-higher standards for data and code quality and sharing while continuing to push boundaries with our scientific questions and, often, advancing field and laboratory research programs. We argue that the scope of questions enabled by the open science movement has far outpaced researchers' skills in being able to answer them, and we need more supported time to learn and teach these skills. Indeed, limited time and skills explain why many researchers do not participate fully in open science practices (Gomes et al. 2022, O'Dea et al. 2021). Skill building is important for early-career researchers, who are central to changing norms

in science (Gownaris et al. 2022). It is also important for researchers at every career stage, who need continued learning time as part of their jobs so they can continue to participate in modern open science practices throughout their careers, whether as researchers, mentors, or supervisors (Robinson & Lowndes 2022).

Paid learning time is something that needs to be built into jobs across career stages and is an issue of diversity, equity, and inclusion: Continuing the trend that scientists learn to code on their own time exacerbates societal inequities. Open educational resources exist [e.g., STAT 545 (<https://stat545.com>) and Data Carpentry for Biologists (<https://datacarpentry.org/semester-biology>)]; what is needed is paid time and career incentives for teaching and skill building. Given the rapid pace of progress in open science tools and the current lack of institutional incentives to engage with them (Soeharjono & Roche 2021), we strongly advocate for giving working marine scientists opportunities to be paid to learn regularly, including through paid leave or sabbaticals.

In our experience, the conversation has shifted from “I don’t want to do open science” to “I don’t know how” and “I don’t have time.” This shift deserves a celebration of the long-term work of the global open science movement, such that the motivation and benefits of open science are understood by researchers. Now, what is missing is how researchers can learn the skills to reap those benefits while continuing their disciplinary work and achieving a work–life balance (Hostler 2023). Even if open science practices will eventually make researchers more efficient, the learning curve may require an unacceptably high investment of time and effort that is unfeasible without paid time across career stages. It is especially crucial to reckon with the way open science is adding strain to researchers’ workloads in the context of the burnout crisis in academia that was accelerated by the COVID-19 pandemic (Gewin 2021).

In this article, we discussed some trends in the open science movement—particularly in the context of data, code, publishing, and community science—that marine scientists can engage in. We are grateful for the academic publishers and governments that are spearheading policies to require open science practices, as well as groups changing tenure and promotion structures to reward not only high-impact lead-author publications but all types of open science contributions, such as software packages, data, open educational resources, and artwork [e.g., the Scholarly Publishing and Academic Resources Coalition (SPARC; <https://sparcopen.org>) and the National Academies of Science, Engineering, and Medicine’s Roundtable on Aligning Incentives for Open Scholarship (Natl. Acad. Sci. Eng. Med. 2023)]. And we call for more support for researchers to develop the skills they need to meet those policies: Transforming how we are educated so that marine scientists learn data science tools and open science practices as part of their coursework, rather than solely via voluntary participation in extracurricular groups, would lift a huge burden off of researchers at all career stages. Since many university faculty are currently ill-equipped to teach these courses (Emery et al. 2021), institutional incentives to support teachers and learners of open science skills—and to hire open data science professors of practice—are key. Institutional support for open science communities of practice like R-Ladies and EcoDataScience can also provide a support system and structured learning environment for scientists at all career stages.

Our career paths provide examples of what the open data science vision can look like in marine science. Alexa Fredston benefited enormously from colleagues whose employment was secure enough that they could invest substantial time and energy in maintaining their own cutting-edge skills and teaching her and others. Since then, she has been able to pay it forward by becoming a teacher of these same methods and ideas—for example, by teaching others how to use GitHub and speaking to audiences of programmers about environmental applications. Through open science, Julia Lowndes has engaged with global efforts supporting science and scientists. Through Openscapes, her work helps researchers in academia, government, and nonprofits explore these resources together with their teams and a cohort of peers, developing habits for collaboration and

reducing the loneliness that is such an unfortunate common feeling when learning to code alone. Together, we have both conducted marine synthesis research that relied on others making data publicly available along with associated metadata so that the data could be reused accurately. Most important of all, we have both been part of marine data science communities that celebrated our successes, normalized our failures, and opened up pathways for us to lead. We recognize that, to date, this experience has been relatively rare in the marine sciences. Our hope for this article is to provide a welcome to marine scientists so that everyone in our field feels empowered and included with opportunities to join the open science movement.

## DISCLOSURE STATEMENT

A.L.F. is treasurer and on the Board of Directors of the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology (SORTEE). This is a volunteer role with no compensation. SORTEE is a registered 501(c)3 nonprofit organization in the United States.

## ACKNOWLEDGMENTS

Open data science work is highly collaborative, and we are grateful to the many people and communities supporting research and research culture. We are grateful to Dominique Roche, Rachel Turba, and Amanda Whitmire for providing invaluable feedback on earlier drafts of this article, and to the NCEAS, EcoDataScience, R, and SORTEE communities. Additionally, J.S.S.L. would like to thank Erin Robinson and the Openscapes and Mozilla communities.

## LITERATURE CITED

- Ancion Z, Borrell-Damián L, Mounier P, Rooryck J, Saenen B. 2022. *Action plan for Diamond Open Access*. Rep., Sci. Eur., Brussels, Belg.; cOAlition S, Eur. Sci. Found., Strasbourg, Fr.; Open Sch. Commun. Eur. Res. Area Soc. Sci. Humanit. (OPERAS), Brussels, Belg.; and Fr. Natl. Res. Agency, Paris, Fr.
- Barker M, Chue Hong NP, Katz DS, Lamprecht A-L, Martinez-Ortiz C, et al. 2022. Introducing the FAIR Principles for research software. *Sci. Data* 9(1):622
- Barone L, Williams J, Micklos D. 2017. Unmet needs for analyzing biological big data: a survey of 704 NSF principal investigators. *PLoS Comput. Biol.* 13(10):e1005755
- Barros C, Luo Y, Chubaty AM, Eddy IMS, Micheletti T, et al. 2023. Empowering ecological modellers with a PERFICT workflow: seamlessly linking data, parameterisation, prediction, validation and visualisation. *Methods Ecol. Evol.* 14(1):173–88
- Bastille K, Hardison S, deWitt L, Brown J, Samhoury J, et al. 2021. Improving the IEA approach using principles of open data science. *Coast. Manag.* 49(1):72–89
- Berberi I, Roche DG. 2022. No evidence that mandatory open data policies increase error correction. *Nat. Ecol. Evol.* 6(11):1630–33
- Binley AD, Bennett JR. 2023. The data double standard. *Methods Ecol. Evol.* 14(6):1389–97
- Binley AD, Vincent JG, Rytwinski T, Proctor CA, Urness ES, et al. 2023. Patterns of community science data use in peer-reviewed research on biodiversity. *Biol. Conserv.* 280:109985
- Blasco GD, Ferraro DM, Cottrell RS, Halpern BS, Froehlich HE. 2020. Substantial gaps in the current fisheries data landscape. *Front. Mar. Sci.* 7:612831
- Bledsoe EK, Burant JB, Higino GT, Roche DG, Binning SA, et al. 2022. Data rescue: saving environmental data from extinction. *Proc. R. Soc. B* 289(1979):20220938
- Braga PHP, Hébert K, Hudgins EJ, Scott ER, Edwards BPM, et al. 2023. Not just for programmers: how GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods Ecol. Evol.* 14(6):1364–80
- Burton C, Duran G, Wright V, Chmiel R. 2023. Strategies for and barriers to collaboratively developing anti-racist policies and resources as described by geoscientists of color participating in the Unlearning Racism in Geoscience (URGE) program. *Earth's Future* 11(2):e2022EF002957

- Butler L-A, Matthias L, Simard M-A, Mongeon P, Haustein S. 2023. The oligopoly's shift to open access. How for-profit publishers benefit from article processing charges. Zenodo 7806651. <https://doi.org/10.5281/zenodo.7806651>
- Callaghan CT, Mesaglio T, Ascher JS, Brooks TM, Cabras AA, et al. 2022. The benefits of contributing to the citizen science platform iNaturalist as an identifier. *PLoS Biol.* 20(11):e3001843
- Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Sci. J.* 19(1):43
- Champion C, Hobday AJ, Tracey SR, Pecl GT. 2018. Rapid shifts in distribution and high-latitude persistence of oceanographic habitat revealed using citizen science data from a climate change hotspot. *Glob. Change Biol.* 24(11):5440–53
- Chytrý M, Pillar VD, Price JN, Wagner V, Wisser SK, Zelený D. 2023. The benefits of publishing in society-owned scientific journals. *Appl. Veg. Sci.* 26(1):e12705
- Cooke SJ, Nguyen VM, Young N, Reid AJ, Roche DG, et al. 2021. Contemporary authorship guidelines fail to recognize diverse contributions in conservation science research. *Ecol. Solut. Evid.* 2(2):e12060
- Culina A, Baglioni M, Crowther TW, Visser ME, Woutersen-Windhouver S, Manghi P. 2018. Navigating the unfolding open data landscape in ecology and evolution. *Nat. Ecol. Evol.* 2(3):420–26
- Czapanskiy MF, Ponganis PJ, Fahlbusch JA, Goldbogen JA. 2022. Compendium of R code and data for “An accelerometer-derived ballistocardiogram method for detecting heartbeats in free-ranging marine mammals.” *GitHub*. <https://github.com/FlukeAndFeather/cetaceanbcg>
- Darwin Core Task Group. 2009. Darwin Core. *Biodiversity Information Standards*. <http://www.tdwg.org/standards/450>
- Emery NC, Crispo E, Supp SR, Farrell KJ, Kerkhoff AJ, et al. 2021. Data science in undergraduate life science education: a need for instructor skills training. *BioScience* 71(12):1274–87
- Eur. Sci. Found. 2023. What is cOAlition S? *European Science Foundation*. <https://www.coalition-s.org/about>
- Fanelli D. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90(3):891–904
- Fenwick I, Davis A, Williams A, Butland S, Lowndes JSS. 2023. Community building for Black environmental and marine researchers: Pathways to Open Science. *Openscapes*, Apr. 26. <https://www.openscapes.org/blog/2023/04/26/pathways-report>
- Filazzola A, Lortie C. 2022. A call for clean code to effectively communicate science. *Methods Ecol. Evol.* 13(10):2119–28
- Fu DY, Hughey JJ. 2019. Releasing a preprint is associated with more attention and citations for the peer-reviewed article. *eLife* 8:e52646
- Gaynor KM, Azevedo T, Boyajian C, Brun J, Budden AE, et al. 2022. Ten simple rules to cultivate belonging in collaborative data science research teams. *PLoS Comput. Biol.* 18(11):e1010567
- Geiger RS. 2018. Reports from the BIDS Best Practices in Data Science Series. *StuartGeiger.com*, Nov. 11. <https://newweb.stuartgeiger.com/articles/2018-11-11-BIDS-best-practices>
- Genome Stand. Consort. 2023. GSC Minimum Information about any Sequence (MIxS). *Genome Standards Consortium*. <https://www.gensc.org/pages/standards-intro.html>
- Gewin V. 2021. Pandemic burnout is rampant in academia. *Nature* 591(7850):489–91
- Gil Y, David CH, Demir I, Essawy BT, Fulweiler RW, et al. 2016. Toward the geoscience paper of the future: best practices for documenting and sharing research from data to software to provenance. *Earth Space Sci.* 3(10):388–415
- Gomes DGE, Pottier P, Crystal-Ornelas R, Hudgins EJ, Foroughirad V, et al. 2022. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proc. R. Soc. B* 289(1987):20221113
- Gownaris NJ, Vermeir K, Bittner M-I, Gunawardena L, Kaur-Ghumaan S, et al. 2022. Barriers to full participation in the open science life cycle among early career researchers. *Data Sci. J.* 21(1):2
- Halpern BS, Berlow E, Williams R, Borer ET, Davis FW, et al. 2020. Ecological synthesis and its role in advancing knowledge. *BioScience* 70(11):1005–14
- Hipsley CA, Sherratt E. 2019. Psychology, not technology, is our biggest challenge to open digital morphology data. *Sci. Data* 6(1):41
- Hörstmann C, Buttigieg PL, Simpson P, Pearlman J, Waite AM. 2021. Perspectives on documenting methods to create ocean best practices. *Front. Mar. Sci.* 7:1260

- Hostler TJ. 2023. The invisible workload of open research. *J. Trial Error*. <https://doi.org/10.36850/mr5>
- Jenkins GB, Beckerman AP, Bellard C, Benítez-López A, Ellison AM, et al. 2023. Reproducibility in ecology and evolution: minimum standards for data and code. *Ecol. Evol.* 13(5):e9961
- Kelly R, Fleming A, Pecl GT. 2019. Citizen science and social licence: improving perceptions and connecting marine user groups. *Ocean Coast. Manag.* 178:104855
- Leonelli S, Spichinger D, Prainsack B. 2015. Sticks *and* carrots: encouraging open science at its source. *Geogr. Environ.* 2(1):12–16
- Lowndes JSS. 2019. Open software means kinder science. *Observations*, Dec. 10. <https://blogs.scientificamerican.com/observations/open-software-means-kinder-science>
- Lowndes JSS, Best BD, Scarborough C, Afflerbach JC, Frazier MR, et al. 2017. Our path to better science in less time using open data science tools. *Nat. Ecol. Evol.* 1(6):0160
- Lowndes JSS, Froehlich HE, Horst A, Jayasundara N, Pinsky ML, et al. 2019. Supercharge your research: a ten-week plan for open data science. *Nature*. <https://doi.org/10.1038/d41586-019-03335-4>
- Lowndes JSS, Robinson EM. 2021. The NASA-Openscapes Framework. Zenodo 5090115. <https://doi.org/10.5281/zenodo.5090115>
- Lowndes JSS, Robinson EM. 2022. *Supporting open science as a daily practice*. Keynote panel presented at the American Geophysical Union Fall Meeting, Chicago, Dec. 12–16
- Marín-Spiotta E, Barnes RT, Berhe AA, Hastings MG, Mattheis A, et al. 2020. Hostile climates are barriers to diversifying the geosciences. *Adv. Geosci.* 53:117–27
- Maureaud AA, Palacios-Abrantes J, Kitchel Z, Mannocci L, Pinsky M, et al. 2023. FishGlob\_data: an integrated database of fish biodiversity sampled with scientific bottom-trawl surveys. OSF Preprints 2bcjw. <https://doi.org/10.31219/osf.io/2bcjw>
- McKiernan EC, Bourne PE, Brown CT, Buck S, Kenall A, et al. 2016. How open science helps researchers succeed. *eLife* 5:e16800
- Mekonnen A, Downs C, Effiom EO, Razafindratsima O, Stenseth NC, Chapman CA. 2021. What costs half a year's pay for African scholars? Open access. *Nature* 596(7871):189
- Merow C, Boyle B, Enquist BJ, Feng X, Kass JM, et al. 2023. Better incentives are needed to reward academic software development. *Nat. Ecol. Evol.* 7(5):626–27
- Nakagawa S, Ivimey-Cook ER, Grainger MJ, O'Dea RE, Burke S, et al. 2023. Method Reporting with Initials for Transparency (MeRIT) promotes more granularity and accountability for author contributions. *Nat. Commun.* 14(1):1788
- NASA (Natl. Aeronaut. Space Adm.). 2022. 2023: Year of Open Science. *NASA Transform to Open Science (TOPS)*. <https://nasa.github.io/Transform-to-Open-Science/year-of-open-science>
- Natl. Acad. Sci. Eng. Med. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: Natl. Acad. Press
- Natl. Acad. Sci. Eng. Med. 2023. Roundtable on Aligning Incentives for Open Scholarship. *National Academies of Science, Engineering, and Medicine*. <https://www.nationalacademies.org/our-work/roundtable-on-aligning-incentives-for-open-science>
- Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, et al. 2015. Promoting an open research culture. *Science* 348(6242):1422–25
- NumFOCUS. 2023. Case study: first photograph of a black hole. *NumFOCUS*. <https://numfocus.org/case-studies/first-photograph-black-hole>
- O'Dea RE, Parker TH, Chee YE, Culina A, Drobniak SM, et al. 2021. Towards open, reliable, and transparent ecology and evolutionary biology. *BMC Biol.* 19(1):68
- Ovando D, Caselle JE, Costello C, Deschenes O, Gaines SD, et al. 2021. Assessing the population-level conservation effects of marine protected areas. *Conserv. Biol.* 35(6):1861–70
- Pecl GT, Stuart-Smith J, Walsh P, Bray DJ, Kusetic M, et al. 2019. Redmap Australia: challenges and successes with a large-scale citizen science-based approach to ecological monitoring and community engagement on climate change. *Front. Mar. Sci.* 6:349
- Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, et al. 2015. Open science resources for the discovery and analysis of *Tara* Oceans data. *Sci. Data* 2(1):150023
- Poisot T, Bruneau A, Gonzalez A, Gravel D, Peres-Neto P. 2019. Ecological data should not be so hard to find and reuse. *Trends Ecol. Evol.* 34(6):494–96

- Primack RB, Miller TK, Terry C, Marín-Spiotta E, Templer PH, et al. 2023. Historically excluded groups in ecology are undervalued and poorly treated. *Front. Ecol. Environ.* <https://doi.org/10.1002/fee.2613>
- Ram K. 2013. Git can facilitate greater reproducibility and increased transparency in science. *Source Code Biol. Med.* 8:7
- Ramachandran R, Bugbee K, Murphy K. 2021. From open data to open science. *Earth Space Sci.* 8(5):e2020EA001562
- Reichman OJ, Jones MB, Schildhauer MP. 2011. Challenges and opportunities of open data in ecology. *Science* 331(6018):703–5
- Robinson E, Lowndes JSS. 2022. The Openscapes Flywheel: a framework for managers to facilitate and scale inclusive Open science practices. EarthArXiv X5CQ02. <https://doi.org/10.31223/X5CQ02>
- Roche DG, O’Dea RE, Kerr KA, Rytwinski T, Schuster R, et al. 2022a. Closing the knowledge-action gap in conservation with open science. *Conserv. Biol.* 36(3):e13835
- Roche DG, Raby GD, Norin T, Ern R, Scheuffele H, et al. 2022b. Paths towards greater consensus building in experimental biology. *J. Exp. Biol.* 225(Suppl. 1):jeb243559
- Secchi Disk Seafarers, Lavender S, Beaugrand G, Outram N, Barlow N, et al. 2017. Seafarer citizen scientist ocean transparency data as a resource for phytoplankton and climate research. *PLOS ONE* 12(12):e0186092
- Shea MM, Kuppermann J, Rogers MP, Smith DS, Edwards P, Boehm AB. 2023. Systematic review of marine environmental DNA metabarcoding studies: toward best practices for data usability and accessibility. *PeerJ* 11:e14993
- Soeharjono S, Roche DG. 2021. Reported individual costs and benefits of sharing open data among canadian academic faculty in ecology and evolution. *BioScience* 71(7):750–56
- Stoudt S, Vásquez VN, Martinez CC. 2021. Principles for data analysis workflows. *PLOS Comput. Biol.* 17(3):e1008770
- Sullivan BL, Phillips T, Dayer AA, Wood CL, Farnsworth A, et al. 2017. Using open access observational data for conservation action: a case study for birds. *Biol. Conserv.* 208:5–14
- Walter P, Mullins D. 2019. From symbiont to parasite: the evolution of for-profit science publishing. *Mol. Biol. Cell* 30(20):2537–42
- Watson C. 2022. Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. *Nat. Med.* 28(1):2–5
- White House Off. Sci. Technol. Policy. 2022. *OSTP issues guidance to make federally funded research freely available without delay*. Press Release, White House Off. Sci. Technol. Policy, Aug. 25. <https://www.whitehouse.gov/ostp/news-updates/2022/08/25/ostp-issues-guidance-to-make-federally-funded-research-freely-available-without-delay>
- White House Off. Sci. Technol. Policy. 2023. *Fact sheet: Biden-Harris Administration announces new actions to advance open and equitable research*. Press Release, White House Off. Sci. Technol. Policy, Jan. 11. <https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research>
- Williams JJ, Drew JC, Galindo-Gonzalez S, Robic S, Dinsdale E, et al. 2019. Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLOS ONE* 14(11):e0224288
- Williams JJ, Tractenberg RE, Batut B, Becker E. 2023. Optimizing short-format training: an international consensus on effective, inclusive, and career-spanning professional development in the life sciences and beyond. bioRxiv 2023.03.10.531570. <https://doi.org/10.1101/2023.03.10.531570>
- Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK. 2017. Good enough practices in scientific computing. *PLOS Comput. Biol.* 13(6):e1005510
- Wolkovich EM, Regetz J, O’Connor MI. 2012. Advances in global change research require open science by individual researchers. *Glob. Change Biol.* 18(7):2102–10
- Wong APS, Wijffels SE, Riser SC, Pouliquen S, Hosoda S, et al. 2020. Argo data 1999–2019: two million temperature-salinity profiles and subsurface velocity observations from a global array of profiling floats. *Front. Mar. Sci.* 7:700
- Zastrow M. 2020. Open science takes on the coronavirus pandemic. *Nature* 581(7806):109–10